

立法院第 11 屆第 5 會期

教育及文化委員會第 8 次全體委員會議

AI 主權時代：學門建構與應用、
研究應用倫理及數位透明度之
配套措施

(專題報告)

報告機關：中 央 研 究 院

報告日期：115 年 4 月 29 日

主席、各位委員女士、先生：

今日 大院第 11 屆第 5 會期教育及文化委員會召開全體委員會議，本院承邀列席報告，謹就「AI 主權時代：學門建構與應用、研究應用倫理及數位透明度之配套措施」進行專題報告。

隨著全球人工智慧（AI）技術的爆炸性成長，臺灣正面臨維護科技自主與文化主權的關鍵時刻，本院亦積極投入「主權 AI」的建構，並致力在學術研究倫理與數位透明度上，持續研擬並落實適切的規範與配套指引。

以下將從「學門建構與應用及主權 AI」、「AI 應用於學術研究之倫理議題」及「數位透明度之配套措施」三個面向進行報告與說明。敬請各位委員指教：

壹、學門建構與應用及主權 AI：

為落實學門建構與應用，本院已成立「AI 推動辦公室」，匯聚院內各領域專業人才，全面推動院內智慧化工作與尖端研究，其具體工作包括支援院內各行政單位的智慧化作業、推動 AI 知能培力教育、提供「AI 合作社」諮詢服務、執行「研發 AI 在人文與科學研究的創新應用」計畫，並持續推動

「中央研究院生成式 AI 風險研究小組」針對最新議題的深入討論。以本院推動「研發 AI 在人文與科學研究的創新應用」計畫為例，支援跨域合作及高運算需求，115 年度預算已編列 1.2 億元。該計畫補助多項前瞻性的研究，例如運用人工智慧解鎖歷史文本，將多版式中文期刊、多語種手寫古文書與日文戶口調查簿進行全文數位化。這些豐富的數位化學術資源，日後不僅推動人文學科的數位轉型，更可作為充實臺灣主權 AI 發展時必備的正確訓練語料。

主權 AI 涵蓋基礎設施、資料、人才與商業網絡四大要素，並須配合監管與倫理框架，我國尤應在能源受限下，發展節能運算或小資料模型。在 AI 主權時代，臺灣所需要的不只是可用的語言模型，更是能夠理解臺灣場景、支援臺灣產業、吻合臺灣價值，並由臺灣自主掌握的基礎模型能力。過去幾年，本院參與 TAIDE(Trustworthy AI Dialogue Engine, 可信任生成式 AI 對話引擎)計畫的發展，正是希望為臺灣建立這樣的基礎，從語言模型進一步延伸到多模態視覺與影片大型模型，形成可支援多元研究與產業應用的共通 AI 底座。

我們特別重視多模態大型語言模型的發展，因為大量真實世界的資訊並不只存在於文字，而是存在於影像、影片、時序事件與行為互動之中。若臺灣能建立具備一般行為分析

能力的多模態基礎模型，便可進一步支援智慧製造、智慧醫療、交通監測、公共安全、長照照護、運動分析、教育科技、社會科學研究與數位人文等多種應用場域。這樣的模型不必為單一任務而生，而是作為可泛化、可轉用、可持續微調的基礎能力平台，讓各產業與研究領域能在其上快速發展終端應用。基礎模型的定義是：具有可延展性、必須是開源碼，最好是能成為由臺灣自行開發、並且是在世界領先群的技術。具有類似條件，由臺灣自行開發的影像技術有 YoLov7, YoLov9，及新近開發的 Taiwan GBA (Taiwan general behavior analyzer)。透過這些多模態基礎模型的整合，很多實用的下游應用系統，例如居家周圍的安控系統，醫療體系的病人照護系統，以及長者的居家照護系統，都能輕易的發展並組合起來。

因此，我們認為，AI 主權的核心不只是模型自主，更是建立屬於臺灣的資料、模型、評測、倫理與應用生態系。以多模態視覺／影音大型語言模型為核心的發展方向，能使臺灣在面對未來 AI 競爭時，不僅擁有技術能力，也擁有理解自身環境、支撐自身產業與服務自身社會的能力。這正是主權 AI 作為臺灣 AI 應用基礎模型的重要意義。

貳、AI 應用於學術研究之倫理議題：

AI 本身並非問題，關鍵在於人類如何選擇與運用該技術，「主權 AI」強調國家對 AI 基礎設施、資料與關鍵技術的掌握，同時包含 AI 應用契合本國需求與符合本國規範要求的確保。因此，建立完善的 AI 管制框架與倫理規範，也是一國 AI 主權的展現，而確立 AI 應用於學術研究之倫理準則及行為規範，更是實踐主權 AI 不可或缺的一環。

人類進入全面受 AI 影響的社會，AI 成為學術研究的課題，同時也以各種方式被應用在各學科領域中，並且多方面地改變了學術研究的過程與產出。AI 對於學術發展的無窮機會與潛在風險，需要分析與評估，特別是生成文字與圖像的生成式 AI。

AI 應用於學術研究的倫理問題主要有兩個面向，均有待處理與規範：

- 一、研究倫理：重點在保護研究對象，例如 Chatbot 可以作為線上訪談、問卷等質性研究的工具，但可能會影響到研究對象的資訊隱私或自主性，特別是研究對象如果是易受傷害族群。
- 二、學術倫理：重點是維護研究誠信。例如 AI 可以用來收集、彙整研究資料、分析問題、編輯文字、生

成文本、潤稿或校稿、轉換語言等等，但同時也可能發生研究人員以譯代稿、以生成文本代替研究著作、虛構註解或參考文獻等學術倫理問題。

本院除了在「中央研究院研究人員及研究技術人員倫理規約」中明定本院人員進行研究工作時，不得有研究舞弊、造假、學術論著抄襲或其他違反各專業領域研究倫理規範之行為外，特別發布「中央研究院使用生成式 AI 之參考指引」，協助研究人員在運用生成式 AI 時，做出正確判斷，實踐負責任的研究行為。該指引明確揭示三大核心準則：首先是「確保 AI 生成內容的正確性」，研究人員必須進行確實的人工檢核；其次為「維護個人隱私與資訊安全」，避免將未公開、機密或個人資料提供給生成式 AI；最後是「實踐研究透明性與課責性」，若實質使用生成式 AI，均應依照學術慣例予以揭露並正確註明，研究人員須對自身研究行為及成果負完全責任。並將其落實於研究管理流程中，包含於徵求說明文件中明確要求研究人員在撰寫與執行研究計畫時如有使用生成式 AI 之協助，均應對其正確與正當性負完全責任，以確保科研成果之誠信，該指引置於本院計畫申請系統中，供申請人於申請階段即可隨時下載閱覽。在此之外，本院「生成式 AI 風險研究小組」對於 AI 應用於學術研究的倫理問題及

其對策，也進行了初步研究（見附件）。

針對運用 AI 從事學術研究及發表所生研究倫理與學術倫理問題，目前面臨的課題主要是，除了訂定「參考指引」等建議性的規範外，如何解釋及適用既有的相關倫理規範，例如在指引已要求落實揭露的基礎上，使用 AI 生成的文章是否應該揭露？內容大量使用 AI 生成的研究成果是否構成「抄襲」或「造假」？相關機關是否需要針對 AI 訂定新的行為規範，以供遵循？從學術自由及科學發展的角度來說，使用 AI 可以促進學術的發展與學術品質的精進，應該正面看待。至於使用 AI 對於學術研究所生的風險與問題，除了靠個別研究人員的自律外，需要由各學術社群針對學科特性逐步累積經驗，形成判準。總之，研究人員學術責任與研究誠信的強化，有賴學術界集思廣益、匯集共識，訂出兼顧學術自由與學術責任的倫理規範。

參、數位透明度之配套措施：

因應 AI 主權時代，確保數位透明度與完善的治理配套措施至關重要。本院在此提出以下具體配套作為：

- 一、生成式 AI 風險治理對策：本院「生成式 AI 風險研究小組」已針對生成式 AI 可能帶來的風險，提出五大層面的治理對策與建言，包含：著作權等智慧

財產權保護、個人資料保護與資料治理、技術風險治理與使用者保護、民主影響評估與公共領域風險治理，以及社會衝擊與環境影響之風險治理。

二、語料庫授權機制與資料透明度：充實具臺灣在地視角的主權 AI 語料庫是當務之急。本院正與國研院 TAIDE 計畫密切合作，盤點院內自主研究計畫之研究資料，研議透過「打包授權」或「共同授權」等機制，將優質研究資料導入大型語言模型訓練。同時建立資料使用紀錄與後續更新管理機制，確保訓練資料的透明度與合法性。

三、確保本土知識正確性與防範數位殖民：主權 AI 的核心在於確保本土知識的正確性。應著力於以專業領域知識建構的資料所發展出具備專業領域及文化特色的開源模型，開放提供給廣大的民眾使用及推廣商用。例如本院穩定推動「建構原住民族教育文化知識體系計畫」，將研究與典藏成果進行數位盤點並彙整為目錄。此外，透過數位博物館等平台開放平埔原住民族等多元典藏資料。這些多元且正確的在地研究資源，將是充實臺灣主權 AI 語料庫、維護文化主體性與數位透明度的重要防線。

肆、結語：

綜合上述說明，「主權 AI」發端於人類思想，觀念與實踐仍在持續發展中。對於臺灣的意義在於我們必須能夠自主掌控並形塑數位資源及 AI 基礎設施，包括數位軟硬體、服務、技術及人才等，避免在關鍵層面上過度依賴外力，同時必須確保 AI 系統的安全性，完善資料保護的法律規範，維護人權保障與民主憲政的核心價值，並將文化自主性與創造力視為主權的重要內涵，在強化權利保護與提升科技競爭力之間持續尋求平衡。

以上報告，敬請

主席及各位委員、女士、先生指教，謝謝！

AI 研究應用倫理

問題與對策

壹、AI 與研究：問題面向.....	10
一、研究倫理：保護研究對象.....	10
二、學術倫理：維護研究誠信.....	10
貳、倫理與規範.....	10
一、研究倫理：Set of Professional Values.....	10
二、學術倫理：Good Scientific Practice.....	10
參、思考與觀點：機會與風險.....	12
一、研究倫理.....	12
二、學術倫理：lege artis.....	12
(一) 思考層次.....	12
(二) 權衡因素.....	13
三、研究倫理與學術倫理的共通課題：個資保護.....	13

壹、AI 與研究：問題面向

一、研究倫理：保護研究對象

- 訪談、民意調查、人體試驗…

問題：→ 以 AI 作為質性研究的工具：Chatbots、精準醫療技術…

二、學術倫理：維護研究誠信

- 收集資料、彙整資料、問題分析、議題設定、生成觀點、文字編輯、文本生成、文字校對、潤稿、語言轉換（翻譯）…

問題：→ 以譯代稿、以生成文本代稿、虛構註解、虛構參考文獻…

貳、倫理與規範

一、研究倫理：Set of Professional Values

- 人體研究法
- 人體研究計畫諮詢取得原住民族同意與約定商業利益及其應用辦法
- 個人資料保護法
- 人體生物資料庫管理條例
- 衛福部函釋
- 各 IRB 的現階段共識
- …

二、學術倫理：Good Scientific Practice

中研院

- 中央研究院研究人員及研究技術人員倫理規約
第 6 條：「本院人員進行研究工作時，應以有效及安全之方式完成學術成果，遵守本院有關研究活動之規定，並不得有研究舞弊、造假、學術論著抄襲或其他違反各專業領域研究倫理規範之行為。」
- 中央研究院使用生成式 AI 之參考指引

國科會

- 國家科學及技術委員會學術倫理案件處理及審議要點

教育部

- 教育部專科以上學校學術倫理案件處理原則

各大學

- 國立成功大學 (2025 年 1 月)。國立成功大學生成式 AI 於教學研究的學術誠信指引。 <https://oai.web2.ncku.edu.tw/p/406-1072-215366,r154.php?Lang=zh-tw>
- 國立陽明交通大學 (2023 年 5 月)。國立陽明交通大學生成式 AI 用於學術研究之參考指引。
<https://oaeri.nycu.edu.tw/userfiles/oaerich/files/20231006111309801.pdf>

各學會

- 臺灣學術倫理教育學會 (2023 年 2 月 23 日)。人工智慧技術對學術倫理的影響及因應建議。
https://www.taaee.org.tw/docs/20230223_conclusion_final.pdf

國際組織

- Committee on Publication Ethics (COPE) Council. (2023, February 13). COPE position – Authorship and AI.
<https://doi.org/10.24318/cCVRZBms>
- Empfehlung zur Sicherung der guten wissenschaftlichen Praxis beim Umgang mit Künstlicher Intelligenz Beschlossen durch das Präsidium der Leibniz-Gemeinschaft am 29. November 2024.
https://www.leibniz-gemeinschaft.de/fileadmin/documents/3085-2906/Leibniz_Empfehlung_zur_KI_Beschluss_Praesidium_2024_11_29_.pdf

歐洲

- ALLEA (2023): The European Code of Conduct for Research Integrity. <https://www.alleageneralassembly.org/wp-content/uploads/2023/06/European-Code-of-ConductRevised-Edition-2023.pdf>

德國

- DFG (2023): Stellungnahme des Präsidiums der Deutschen Forschungsgemeinschaft (DFG) zum Einfluss generativer Modelle für die Text- und Bilderstellung auf die

Wissenschaften und das Förderhandeln der DFG.

<https://www.dfg.de/resource/blob/289674/ff57cf46c5ca109cb18533b21fba49bd/230921stellungnahme-praesidium-ki-ai-data.pdf>

- Hochschulforum Digitalisierung (2024): Leitlinien zum Umgang mit generativer KI.

https://hochschulforumdigitalisierung.de/wp-content/uploads/2024/02/HFD_Blickpunkt_KILeitlinien_final.pdf

奧地利

- KI und wissenschaftliches Arbeiten und Schreiben.
<https://studieren.univie.ac.at/umgang-mit-ki-im-studium/ki-und-wissenschaftliches-arbeiten-und-schreiben/>

- 命題：

- ◆ 解釋現行規範？
- ◆ 訂定新的行為規範？
- ◆ 「其他違反學術倫理行為」作為概括條款？

參、思考與觀點：機會與風險

一、研究倫理

四原則：尊重自主原則、不傷害原則、行善原則、公平原則

二、學術倫理：lege artis

(一) 思考層次

- 禁止使用？
Ex. 學術審查
- 容許使用？使用 AI ≠ 違反學術倫理
 - ◆ 揭露義務：是否使用？使用目的？使用程度（範圍）？
 - ◆ Authorship 與「文責」：自然人？
 - ◆ 學術評價：中立原則？違反學術倫理？
 - ◆ ...

(二) 權衡因素

- 學術品質與學術產量
- 學術自由與學術自律
- ...

三、研究倫理與學術倫理的共通課題：個資保護